PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G06F 17/30

A1

(11) International Publication Number: WO 97/40454

(43) International Publication Date: 30 October 1997 (30.10.97)

(21) International Application Number: PCT/IB97/00439

(22) International Filing Date: 25 April 1997 (25.04.97)

(30) Priority Data:
08/637,844 25 April 1996 (25.04.96) US

(71) Applicant: PHILIPS ELECTRONICS N.V. [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(71) Applicant *(for SE only)*: PHILIPS NORDEN AB [SE/SE]; Kottbygatan 7, Kista, S-164 85 Stockholm (SE).

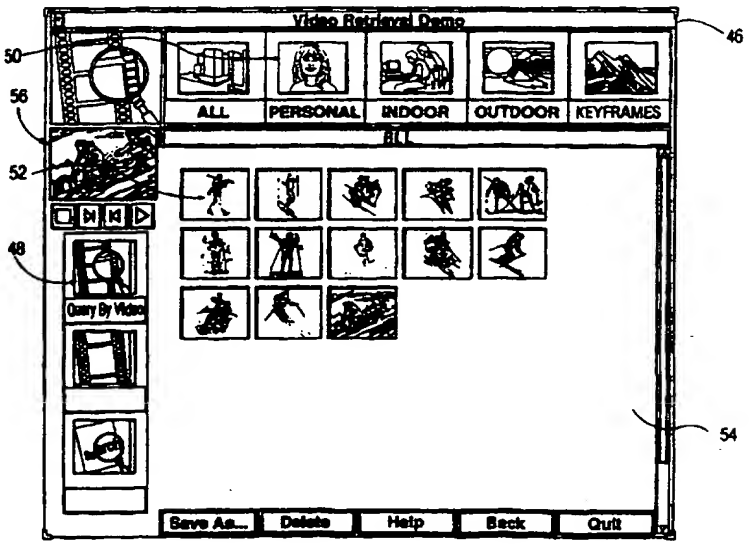
(72) Inventors: DIMITORVA, Nevenka; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). ABDEL-MOTTALEB, Mohamed, S.; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(74) Agent: STRIJLAND, Wilfred; Internationaal Octrooibureau B.V., P.O. Box 220, NL-5600 AE Eindhoven (NL).

(81) Designated States: JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: VIDEO RETRIEVAL OF MPEG COMPRESSED SEQUENCES USING DC AND MOTION SIGNATURES



(57) Abstract

Signatures from MPEG or Motion JPEG encoded video clips are extracted based on the luminance, chrominance, and motion vector values. The signatures are stored in a database, along with corresponding location, size, and time length information. The signature of a query video clip, also encoded in the MPEG or Motion JPEG format, is extracted. The signature of the query video clip is comparred with the signatures stored in the meta database, with video clips having signatures similar to the signature of the query video clip identified. The video clips are then retrieved and displayed by selection of a user.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

Video retrieval of MPEG compressed sequences using DC and motion signatures.

BACKGROUND OF THE INVENTION

The invention relates to storing and retrieving from large digital video archives video images encoded using the Moving Picture Experts Group (MPEG) encoding standard or the Motion JPEG (Joint Photographic Experts Group), and specifically to the extraction, based on DC coefficients and motion vectors, of video sequence signatures from MPEG or Motion JPEG compressed videos, and the search for and retrieval of videos based on the extracted signatures.

The MPEG video compression standard is described in Practical Digital Video With Programming Examples In C, by Phillip E. Mattison, John Wiley and Sons, 1994, chapter 11, pages 373 through 393, and in "MPEG: A Video Compression Standard for Multi-Media Applications", by Didier Le Gall, Communications of the ACM, April 1991, vol. 34, no. 4, pps. 47 through 58, and the JPEG video compression standard is described in "The JPEG Still Picture Compression Standard", by Gregory K. Wallace, Communications of the ACM, April 1991, vol. 34, no. 4, pps. 31 through 44.

Also applicable to video images encoded using the Motion JPEG standard, the present invention is set forth herein with reference to video images encoded using the MPEG standard.

MPEG is used for digitally encoding motion pictures for use in the information processing industry. With this standard, video images can be stored on CD-ROMs, magnetic storage, and in random access memory (RAM and ROM). The MPEG standard allows video images to be transmitted through networks such as ISDNs, wide area networks, local area networks, the INTERNET™, the INTRANET™.

Video clips or video streams are sequences of an arbitrary number of video frames or images. An example of a video clip is images from a television news show. MPEG video clips encoded as MPEG-video or using the MPEG system-layer encoding may have signatures extracted, in accordance with the present invention.

In the MPEG standard, the color representation is YCrCb, a color scheme in which luminance and chrominance are separated. Y is a luminance component of color,

and CrCb are two components of chrominance of color. For each four pixels of luminance, one pixel of Cr and one pixel of Cb is present. In the MPEG standard, the chrominance information is subsampled at one-half the luminance rate in both the horizontal and vertical directions, giving one value of Cr and one value of Cb for each 2X2 block of luminance pixels. Chrominance and luminance pixels are organized into 8X8 pixel blocks (or blocks). Pixel blocks are transformed into the frequency domain using the discrete cosine transform (DCT) operation, resulting in DC and AC components corresponding to the pixel blocks.

In the MPEG standard, images in a sequence are represented by four types: I frame, P frame, B frame, or D frame. Each image is divided into slices, with a slice comprising one or more macro blocks. Slices are typically contiguous macro blocks.

A macro block comprises four 8X8 blocks of luminance pixels and one each 8X8 block of two chrominance (chroma) components. Therefore, a macro block comprises the DCT coefficients for four 8X8 blocks of luminance pixels and one 8X8 block for each of two chrominance coefficient pixels. Alternatively, the macro block may be encoded using forward or backward motion vectors, for B or P frames only. A forward motion vector of a frame is based on motion relative to a previous frame, while a backward motion vector of a frame is based on motion relative to a subsequent frame.

Within an image video clip, the value of a DC coefficient is encoded relative to the previous DC coefficient, with DC values for luminance being encoded relative to other luminance values and DC values for chrominance being encoded relative to chrominance values.

The MPEG standard comprises MPEG-video, MPEG-audio, and MPEG system-layer encoding (which incorporates MPEG-video, MPEG-audio, and information regarding how the two interact), and allows motion video to be manipulated in a cost-effective memo.

Building large video archives which allow video clips to be stored, retrieved, manipulated, and transmitted efficiently necessitates various technologies, such as video analysis, content recognition, video annotation, and browsing. For a user, the most important capability is efficient retrieval based on the content of video clips. The existing methods for content-based retrieval rely principally on the extraction of key frames or on text annotations.

Video browsing systems which rely on text, retrieve video sequences by key word annotation. The textual annotations which are normally stored separately can be indexed using full text retrieval methods or natural language processing methods.

Browsing systems which use key frames to represent video sequences rely on the idea of detecting shot boundaries and choosing certain frames as key frames. A shot is a contiguous number of video frames that convey part of a story. Most modern movies contain over a thousand cuts (a cut is a change between shots), requiring an intelligent video retrieval program to process some thousand frames per movie. To see what is in the video, a user must preview the key frames in the above-mentioned browsing systems.

Further, the above-mentioned browsing systems use individual key frames and motion to search for video clips, without accounting for the sequence of key frames to represent the video clips when submitting the whole video clip as a query.

An alternative retrieval method for browsing is by displaying particular frames of the video sequence, allowing a user to review and select the particular video sequence. This alternative method is time consuming.

SUMMARY OF THE INVENTION

Therefore, an object of the present invention is to retrieve video sequences without relying on text annotations.

A further object is to retrieve video sequences using signatures of representative frames, taking into account motion between frames.

Another object is to extract signatures from video clips digitally-encoded, using the MPEG encoding standard, based on DC components of the DCT coefficients and motion vectors (the signatures extracted by the present invention being referred to as DC+M signatures).

A further object is to retrieve video clips quickly and efficiently from a large database of video footage.

Another object is to retrieve from local or remote databases video clips similar to a query video clip using signatures of the database video clips and the query video clip.

Still another object is archiving video clips of Motion JPEG compressed videos and MPEG compressed videos.

An additional object is to re-encode video clips previously encoded using encoding standards other than Motion JPEG or MPEG into one of the Motion JPEG or MPEG format, and extract and store signatures therefrom.

Another object is a retrieval method that takes a video clip as a query and searches a database for clips with similar content, facilitating retrieval of clips from lengthy

video footage for the purpose of editing, broadcast news retrieval, and copyright violation determination.

In the present invention, video clips encoded using Motion JPEG are considered as MPEG video streams in which all frames are intracoded frames (no P frames or B frames are included, with all frames being I frames). In video clips encoded using the MPEG encoding standard, DC components and motion vectors are determined conventionally. In the case of Motion JPEG, a query clip is a sequence of JPEG frames. In video clips encoded using Motion JPEG, in the present invention, the above-mentioned signatures use only the DC color information (the DC component of the DCT coefficients), without using motion information. The DC color information is determined conventionally.

Further, a signature of a video clip (herein after referred to as a video clip signature) is represented by the sequence of signatures extracted from representative MPEG frames (herein after referred to as frame signatures) within that video clip. Frame signatures comprise DC components and motion vector components corresponding to pairs of windows within each frame. Therefore, the signatures extracted are referred to as the DC+M signatures. In the DC+M signature extraction method, each video clip is represented by a sequence of frame signatures, which requires less storage space than do the encoded video objects or clips. This DC+M method is fast, because the DC coefficients and the motion vectors from MPEG frames can be extracted without fully decoding the MPEG frames. The signatures can be extracted in real time, while the video clips are being encoded into the MPEG format or while the MPEG-encoded video clips are being at least partially decoded.

After signatures of the video clips have been extracted the video retrieval compares a signature of a video clip (a query video clip) against a database storing signatures of video clips (database video clips). Two video clips are compared as sequences of signatures. The comparison is performed for each video clip in the video clip database, and a record or score is kept of the degree of matching or similarity between the signature of the query video clip and the signature of each database video clip. The similarity is based upon the Hamming distance measure between the signatures of the two video clips.

The user can then select which, or all, of the video clips to be viewed according to the score. The order of signatures within the sequence is based on the order of frames in the video clip. The sequential order of signatures is used as a basis to represent the temporal nature of video in the present invention.

These together with other objects and advantages which will be subsequently apparent, reside in the details of the construction and operation as more fully

herein after described and claimed, reference being had to the accompanying drawings forming a part hereof, wherein like numerals referred to like parts throughout, as follows:

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is an overview of the signature extraction;

Figure 2 is an overview of the signature comparison;

Figure 3 is a block diagram of the system architecture for video retrieval;

Figure 4 is a diagram of signature extraction from window pairs in a video frame;

Figure 5 is a diagram showing a macro block structure;

Figure 6 is a diagram explaining derivation of bits in a signature;

Figure 7 is a diagram showing organization of signature bits;

Figures 8(A) and 8(B) are detailed diagrams of mapping windows in images of different sizes;

Figure 9 is a flowchart for mapping windows and calculating DC coefficients;

Figure 10 is a flowchart showing the signature extraction and the archival process;

Figures 11(A), 11(B), 11(C), and 11(D) show examples of a comparison between frames in a database video clip and frames in a query video clip in the retrieval process;

Figure 12 is a display showing a demonstration of query video clips and database video clips retrieved as displayed on a user interface.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

There are two main aspects of the present invention:

(1) archival and signature extraction, and

(2) retrieval of video clips using signatures.

Archival and signature extraction refer to extracting signatures of video clips and storing the extracted signatures of the video sequences in a database. Retrieval refers to extracting the signature from a query video clip (which video clip is one designated by the user for which other, similar video clips are to be identified), then comparing that query video clip signature against the signatures representing the video clips from the database.

A video clip from which a signature is to be extracted must be at least partially encoded using the MPEG or Motion JPEG standard. Such encoding must be at least to the level of having DC coefficients and motion vectors (if the video clip is MPEG-encoded), without quantization, run-length encoding and Huffman coding. In addition, a video clip must be at least partially decoded, meaning that it must be Huffman decoded, run-length decoded, and dequantized, and have DCT coefficients and motion vectors (if the video clip is MPEG-encoded). Motion vectors are present only in video clips at least partially encoded or decoded using the MPEG encoding standard. For video clips at least partially encoded or decoded using the Motion JPEG encoding standard, motion vectors are not present.

Figure 1 shows a flowchart of an overview of extracting signatures of representative frames from a video clip. Referring now to Figure 1, positions and sizes of selected window pairs, which are constant for all of the video clip signature sequences stored in a database, are input in step 100. The exact placement of the window pairs within each frame is predetermined, and is fixed for each frame within the video clip from which the signature is being extracted.

In step 101 a video clip is received and the start and end points of the video clip are determined. Step 102 determines whether the video clip is at least partially encoded using the MPEG or the Motion JPEG encoding standard, is not encoded at all, or is encoded using a different encoding standard. If the video clip is not encoded at all or is encoded using a non-MPEG encoding standard, the video clip is at least partially encoded using the MPEG standard in step 104.

Frames from which signatures are to be extracted are determined in step 103. A signature is not necessarily extracted from all frames.

Once the frames from which a signature is to be extracted are determined, DC coefficients and motion vectors for the selected frames at the selected window positions are extracted in step 105.

In step 106, bits corresponding to the qualitative difference between values of the DC coefficients and bits indicating the qualitative difference between values of the motion vectors (DC+M signatures) between each window in the window pair are determined. In step 107, the signature of the frame is formed by concatenating each of the foregoing DC+M signatures for each window pair within the frame. In step 108, the signature of the video clip (the video clip signature) is represented by the sequence of the signatures of the representative frames.

Once the signature of a video clip is extracted, that signature is stored in a database containing signatures of other video clips. More than one signature may be extracted from a particular video clip, depending upon the placement of the windows within the frames. Therefore, there may be multiple signature databases, each storing signatures having windows placed in the same arrangement and locations as other signatures within that database.

After the signature of a query video clip is extracted, a database storing signatures obtained in the manner prescribed herein is searched to locate video clips similar to the query video clip, based on the signature thereof. The searching method can be adjusted to locate and retrieve video clips with varying degrees of similarity.

Figure 2 shows an overview of determining matches between a signature of a query video clip and signatures of other video clips.

First, in step 201, the query video clip is determined, and is at least partially encoded using the MPEG encoding standard. The signature of the query video clip is extracted according to the process of Figure 1.

In step 202, a signature corresponding to a database video clip is selected for comparison to the signature of the query video clip. Therefore, any representative frame signature stored in the local database may be the first signature selected for comparison.

In step 203, the variable **minscore** is initialized to the length of the frame signature + 1. The length of the frame signature in a preferred embodiment of the present invention is 128; therefore, the value of minscore is initialized to the value of 129, which value is larger than the largest possible value for the Hamming distance measure between the signature sequence extracted from the query video clip and the signature sequence extracted from the database video clip. As shown in step 208, below, minscore stores the value of the lowest total Hamming distance measure between the query video clip signature sequence and the database video clip signature sequence calculated for a given database.

In step 204, a variable is initialized, that stores a total of the Hamming distance measure between representative frame signatures from the query video clip and representative frame signatures from the video clip being compared to the query video clip. Also initialized in step 204 is a variable that counts the number of comparisons made between a frame signature of the query video clip and frame signatures of the database video clip.

In step 205, entries from the query video clip signature that are the frame signatures of the representative frames, are placed into correspondence with entries from the

database video signature. The entries are the frame signatures of the representative frames. Further, the sequence of entries from each video clip signature are collectively referred to as signature sequences for the video clip.

The first entry from the query video clip signature is placed into correspondence with the first entry from the database video clip signature. Further entries from each of the query video clip signature and the database video clip signature are placed into correspondence with each other. The location of the frame signature within the signature sequence is not necessarily the same as the location of the frame, from which the frame signature was extracted, within the video clip.

In step 206 a comparison is performed between an arbitrary frame signature in the query video clip signature sequence and up to three frame signatures in the database video clip signature sequence corresponding to the arbitrary frame signature from the query video clip signature sequence. Preferably, a distance of one frame (which number any number), taking into account any offset or shifting between the entries in the query video clip signature sequence and the database video clip signature sequence, in either the forward or the backward direction in the video clip is permitted for comparison. If a distance greater than one frame in either the forward or backward direction is selected, then a query video clip frame signature may be compared with more than three frame signatures.

In step 207, the Hamming distance measure from the signature of the query video clip to the signature of each corresponding frame in the database video clip is determined for the current iteration of steps 204 through 210. The distance variable is updated accordingly, as is the variable tracking the number of comparisons made. One "score" of the total Hamming distance measure between frames compared for each shift is recorded.

Then, in step 208, if the value of "score" is lower than the value of minscore, the value of "score" is stored in minscore.

The database typically will contain more entries than the query video clip signature: then the signature sequence from the query video clip is compared against each series of entries from the database video clip, preserving the sequence of frames for each video clip. This is done by first placing in correspondence with each other the first entry in each signature sequence and carrying out the comparison. Then, the entire sequence of frame signatures in the query video clip is shifted by one entry, and the comparison sequence is repeated. The shifting repeats until the sequence of frame signatures from the query video clip signature sequence is compared to the corresponding sequences of entries in the

database.

In step 209, whether the query video signature sequence has been compared to each database video signature sequence for the selected database video clip signature is determined. If not, in step 210, the query video signature sequence is shifted by one entry with respect to the database video signature sequence) and control is returned to step 204 for the next iteration.

In step 209, if the query video signature sequence has been compared to each database video signature sequence for the selected database video clip signature, then step 211 is executed.

In step 211, the value stored in minscore is inserted into array Score, which stores entries of minscore corresponding to each database clip.

Whether all database video clips signature sequences stored in the database have been compared to the query video clip signature sequence is determined in step 212.

If not, then in step 213 the next database video clip sequence is selected, and steps 203 through 212 are repeated.

If all database video clips signature sequences have been compared, the process in Figure 2 is complete.

The similarity between the query video clip signature and the database video clip signature is determined as the number of bits in the signature minus the Hamming distance measure between the foregoing video clip signatures. If the Hamming distance is small, then the similarity between the two video clip signatures is high. Therefore, if database video clip signatures are ordered based on the respective similarity to the query video clip signature, then the database video clip signatures are placed into descending order of similarity. A higher degree of similarity between two video clips indicates that the two video clips are closer in appearance.

On the other hand, if database video clip signatures are ordered based on the respective Hamming distance from the query video clip signature, then the database video clip signatures are placed into ascending order of Hamming distance. A lower degree of Hamming distance again indicates closer in appearance.

Preferably, video signature extraction and storage system 8 shown in Figure 3 is implemented on a computer, such as a SUN™ workstation or a Pentium™-based personal computer. The system 8 includes video source 10, video information retrieval system 18, and user interface 32.

Video source 10 may receive video clips from a variety of sources. The video clips could be already digitally encoded in the MPEG or Motion JPEG format and provided by MPEG video server 12. Further, the video clips may be provided from live video, which is provided by live video source 14, or encoded in the MPEG or Motion JPEG format or in a format other than MPEG, which are provided by network source 16.

Each of sources 12, 14, and 16 may interface to other, respective sources providing video clips thereto and which are not shown in Figure 3. A representative source to which MPEG video server 12 interfaces is the INTERNET™, which may store video clips at ftp sites on computers running the UNIX™ operating system along with an **ftp** daemon or at a variety of Web sites running **http** servers awaiting **html** document requests.

Video source 10 provides to a video information retrieval system 18 video clips; if the latter are from MPEG video server 12, no further encoding is necessary.

If live video source 14 provides video clips, the video clips must be compressed using the MPEG encoding standard before the signature sequences are extracted.

If video clips are provided by the network source 16, they may have been encoded in a format other than the MPEG format; subsequently, the video clips must be partially re-encoded using the MPEG encoding standard by, for example, a video bridge, before extraction of the signature sequences may occur. The encoding may use conventional computer hardware and software.

In video information retrieval system 18, archival and signature extraction process 20 extracts the signature of each video clip.

In addition, process 20 stores the extracted DC+M signatures of video clips received from the MPEG video server 12 in meta database 22. Meta database 22 is referred to as a "meta" database because data that describe other data (i.e., signatures of video clips and other descriptive data thereof) are stored therein. Process 20, may be implemented in software on any UNIX ™-based computer, personal computer, or other platform. Process 20 could also be part of an MPEG encoder or MPEG decoder hardware implementation board.

On the other hand, if process 20 receives live video from live video source 14, process 20 compresses or partially compresses in a conventional way the live video using the MPEG encoding standard, extracts the DC+M signature from the compressed MPEG video clips, and stores the extracted signatures in meta database 22. If the network source 16 transmits video clips, then process 20 re-encodes the video clips into the MPEG format, extracts the DC+M signatures from the re-encoded MPEG video clips,

and stores the signatures in the meta database 22.

Along with the extracted signatures of video clips (which signatures comprise frame signatures), archival and signature extraction process 20 stores in meta database 22 other identifying information such as the location at which the corresponding video clip is stored, the size of the video clip in bytes, the time length of the video clip, and the title of the video clip. The frames for which frame signatures are extracted are also referred to as representative frames.

In Figure 3, retrieval subsystem 24 uses the signature extracted from a query video clip to search meta database 22 for signatures of similar video clips. Retrieval subsystem 24 includes: a similarity calculation process 26, an ordering process 28, and a fetch and display process 30.

The similarity calculation process 26 determines the "similarity" between a signature of a query video clip and signatures of the video clips stored in meta database 22. The ordering process 28 determines the ordering of video clips whose signatures are stored in meta database 22. This uses the "similarity" measure.

The fetch and display process 30 contains pointers for displaying video clips whose signatures are stored in meta database 22. If the video clip is stored at a remote Web site on the INTERNET™, fetch and display process 30 tracks the INTERNET™ node of the video clip, and location on the remote file system.

Each of process 26, process 28 and fetch and display process 30 may be a software program, or hardware or firmware.

User interface 32, is front end software, and written using development kits, such as VISUAL C++™ or VISUAL BASIC ™ in which a user can submit a video clip and display search results. An example of a user interface 32 of the present invention is shown Figure 12.

Extraction of the signature sequences is as follows.

Preferably, each frame signature is represented by 128 bits for signatures extracted from video clips which are MPEG-encoded. (For signatures extracted from video clips which are Motion JPEG-encoded, each frame signature is represented by 96 bits, as explained below.) However, the number of bits may be varied by a user, depending upon a resolution or sensitivity desired in the signature of the video clip. In addition, the signature may include 128 bits for every frame of the video clip, for every other frame of the video clip, etc.

A video clip can be considered to be a sequence of video frames, i.e.,

$\{i_o, \ldots, i_n\}$. A video clip can also be represented by a subset of those frames $\{j_o, \ldots, j_n\}$. The representative frames, which are frames for which a signature is extracted, can be selected based on the MPEG frame pattern or using the key frames extracted from scene transitions. Each frame is represented using a signature based on the DC coefficients of window pairs and their motion vectors.

In the present invention, the video clips can be indexed in the following ways:

1. Using I frames as a basis for deriving the DC+M signatures. This method does not require the extraction of key frames. However, the index generated is larger and the retrieval time is longer; or
2. Using the key frames as a basis: In the case of video footage with long scenes, this method generates fewer frames with which to work.

The DC+M signatures are either local DC+M signatures or global DC+M signatures. Local DC+M signatures are also referred to as frame signatures and are signatures derived from a particular frame and neighboring frames, without consideration of the larger context of frames to which the particular frame belongs.

For both the local DC+M signatures and the global DC+M signatures, the DC components of the frame signatures are extracted in the same way. However, the motion bits of the signature differ between the local DC+M signatures and the global DC+M signatures. The motion bits represent whether the motion vector associated with a frame is zero or non-zero. For a local DC+M signature of a representative frame, this applies to motion between the representative frame and frames immediately surrounding the representative frame. However, for global DC+M signature of a representative frame, the motion bits indicate whether the motion vector is zero or non-zero with respect to motion between the representative frame and frames a number of frames away from the representative frame. The motion signature is typically represented by two bits per window pair in local DC+M signatures.

Also, key frame positions can be used as representative frames for which signatures are extracted. Generating signatures depends on the position of the key frame:

1. If a key frame is an I frame: the DC coefficient is taken from the I frame and from the corresponding motion vectors from the following B or P frame - no extra processing is required;
2. If the key frame is a B frame: the reference frames are considered to obtain the DCT coefficients. The DC coefficients are extracted

from respective macro blocks in the previous I or P frame or in a future I or P frame. The motion vectors in the current frame are used to derive the motion vector signature; and

3. For a P frame: signature extraction processing is moved one frame ahead. Most of the luminance and chrominance blocks in the frame ahead will be intracoded (i.e., all information about the macro block in which the blocks are included is contained therein and described by DCT coefficients only, without using motion vectors), simplifying the extraction of the DC coefficients. To extract the DC coefficients, the DC coefficients from the respective macro block of the previous reference frame are set (which is an I frame or P frame). To obtain motion bits of the frame signature, the motion vectors from the next B frame are used. If the next frame is an I frame, then the motion vectors from the closest future B or P frame are used for the motion signature. Therefore, if the key frame is a P frame, the frame signature is extracted from DC coefficients and motion vectors associated with the above-mentioned frames.

If the positions of the key frames are not known in advance, then the signature extraction will use the I frames in the video clip.

Global DC+M signatures are also referred to as frame signatures and are extracted from a frame with respect to a sequence of frames within a video shot or within a frame pattern encoded by the MPEG standard. For global DC+M signatures, the DC component of the signature is extracted in the same manner as in the case of the local DC+M signatures, described in detail below. The motion part of the signature is determined by tracking the macro blocks corresponding to the window pairs until the next I frame is reached or over a subset of the number of frames in the video clip. Then, a qualitative description, for the relative motion of the window pairs is calculated. In this case, the motion signature can be longer than two bits per window pair. The motion signature reflects the relationship between the windows in the window pair over the subset of the number of frames in the video clip.

The signatures in the present invention are derived from relationships between window pairs, as shown in Figure 4. For each image frame 40, a number of window pairs is selected, in which every window pair corresponds to one part of the

signature. Therefore, the number of window pairs determines the length of the signature. Each window of a window pair corresponds to one macro block or region which covers multiple macro blocks in the picture frame.

A macro block in an MPEG frame corresponds to a 16X16 pixel area, as shown in Figure 5. The chrominance and luminance samples are organized into 8X8 pixel blocks. A macro block includes four 8X8 blocks of luminance pixels and one 8X8 block of each of the two chrominance (or chroma) components, as shown in Figure 5.

The signatures of window pairs may be derived as follows:

(a) for a video sequence containing only I frames (such as when encoded using Motion JPEG): for each I frame, 64 bits are derived from the luminance plane, and two sets of 16 bits are derived from each of the two chrominance planes, for a total of 96 bits for the DC components. There is no contribution from motion vectors (for subsequent matching, whilst using exclusively Motion JPEG encoded video clips, only these 96 bits are used);

(b) for a video sequence containing I, B, and P frames (such as for video sequences using MPEG): a 128 bit signature (also referred to as a key) is derived in which 96 bits are derived from the DC coefficients and 32 bits are derived from the motion information available in the MPEG data stream. For subsequent matching, whilst using signatures extracted from I, B, and P frames, all of the foregoing 128 bits are used.

As shown in Figure 6, frame 40 includes three examples of window pairs: w1 and w1', w2 and w2', and w3 and w3'. The positions of window pairs are selected in advance, but are fixed for the entire meta database 22 of signatures and for a signature of a query video clip. A single video clip, accordingly, may have multiple signatures stored in multiple meta databases 22. For example, if one set of signatures, stored in one meta database 22, concentrating on the middle of frames is desired, matching windows are chosen accordingly. On the other hand, if signatures stored in another meta database 22 are meant to correspond to background areas, window positions are chosen accordingly. In the embodiment there are sixteen window pairs for each frame.

From each set of window pairs w1 and w1', w2 and w2', and w3 and w3' in frame 40 shown in Figure 4, a signature 42 is extracted. In Figure 4, signature 42

includes window signature Sw1 from the signature corresponding to window pair w1 and w1'; window signature Sw2 from the signature corresponding to window pair w2 and w2'; and window signature Sw3 from the signature corresponding to window pair w3 and w3'.

The following is an example of how window signatures for each of the foregoing window pairs are determined. For each window, or macro block, in Figure 6, DC coefficients for the luminance blocks and for each of the two chrominance blocks in Figure 5 are extracted. In Figure 5 four blocks in the luminance plane and two blocks in the chrominance planes are used for signature extraction. The DC components of the luminance blocks and chrominance blocks are determined in a conventional manner.

Using the window pair w1 and w1' as an example, as shown in Figure 6, each window in the window pair has six DC coefficients (DCi). In a preferred embodiment six signature bits S1 through S6 are extracted for each window pair based upon the following:

$$Si = 1 \text{ if } |DCi - DCi'| <= threshold \quad (1)$$

$$Si = 0 \text{ if } |DCi - DCi'| > threshold \quad (2)$$

The result of the extraction for each window pair shown in Figure 6 of the six associated DC components is the signature 42 shown in Figure 4. Signature 42 contains, 6 bits for each of signatures Sw1, Sw2, Sw3, corresponding to window pairs w1 and w1', w2 and w2', and w3 and w3'. In addition, if the video clip for which the signature is being extracted is encoded using the MPEG standard, motion bits (not shown in Figure 4) also contribute to the signature 42.

In a preferred embodiment, a bit in signature S_i is calculated using the above-mentioned pair of DC coefficients based upon the above-mentioned equations (1) and (2). However, the signature could also be calculated using other functions of the DCT coefficients of the MPEG-encoded video clip, for example, by adding a DC component to an AC component, and dividing by an arbitrary number.

Having 16 window pairs is preferred because computers most efficiently store binary numbers in groups of 2^n, and unsigned integers are typically stored in 32 bits in many software compilers, such as, many C™ language compilers. There are 6 blocks in a macro block, and 16 bits per block, from which the DC components are derived: 4 luminance blocks X 16 bits = 64 bits, and 1 block for each of 2 chrominance (Cr and Cb) blocks giving 2 X 16 bits = 32 bits for chrominance.

Motion between windows in a window pair may be described in a qualitative manner as follows, using 2 motion bits:

1. Bits are set to 00 if both windows exhibit zero motion (i.e., if the motion vector is zero);
2. Bits are set to 01 if the first window is static (the motion vector is zero) but the second window has moved (the motion vector is non-zero);
3. Bits are set to 10 if the first window has moved (the motion vector is non-zero) but the second window was static (the motion vector is zero); and
4. Bits are set to 11 if both windows exhibit motion (both motion vectors are non-zero).

There is no contribution from motion bits if the video clip is encoded using the Motion JPEG format.

Figure 7 shows an example of a signature, for window pairs of a frame. The signature 44 shown in Figure 7 is 128 bits in length, organized into groups of 8 bits per 16 window pairs. For example, of the first 8 bits in the signature shown in Figure 7, bits L11 through L14 are the above-mentioned luminance bits for window pair w1 and w1'; bits Cr1 and Cb1 are the chrominance bits for window pair w1 and w1'; and bits M11 and M12 are the motion bits (if present) for window pair w1 and w1'. Eight bits for each of window pairs w2 and w2' through w16 and w16' are organized accordingly in Figure 7.

As shown in Figures 8(A) and 8(B), window size and positions are relative to image sizes. In some case a window will cover only one macro block, as shown in Figure 8(A). In other cases, the window will cover more than one macro block and possibly parts of many macro blocks, as shown in Figure 8(B). In the latter case, the DC value is calculated as the weighted sum of the DC values of the macro blocks which are covered by the window. In that case, the weights are the relative areas of the macro blocks covered by the window.

To normalize window size, as above, if a window covers more than one macro block, the window is mapped to a window of standard size in accordance with the process shown in Figure 9.

Figure 9 is a flowchart for mapping windows and calculating DC coefficients in the present invention. In step 301, coordinates of a window defined by Wxi, Wyi, where i equals 1, . . . 4, in the image of the standard size are selected by the user in advance. Also in step 301, the new image size is provided as input. In step 302, each coordinate of the window in the standard image is mapped to correspond to the size of the

new image. In step 303, weights are calculated based on the partial area of the macro block covered by the window as a part of the total area of the window in the new image for each coordinate of the window defined in step 301. Then, in step 304 and as discussed with reference to Figure 8(B), the DC value is calculated as the weighted sum of the DC values of the macro blocks which are covered by the window.

Figure 10 is a flowchart showing frame signature extraction in the signature archival process. In step 401, an index j indicating which window pair in a frame for which a signature is extracted, is initialized to the value "1". In step 402, the process shown in Figure 10 begins for window pairs W_j and W_j', having DC coefficients of the image and motion vectors M_{wj} and M_{wj}'. In step 403, an index i indicating which of the DC components is being calculated is initialized to the value of "1". In step 404, the ith DC coefficient from window, indicated by DCi, is calculated for each window W_j, and the ith DC coefficient, indicated by DCi', is calculated for each window W'_j, in accordance with Figure 6. In step 405, the absolute value of the difference between DCi and DCi' is compared to an arbitrary threshold amount, selected by the user. If the foregoing difference is less than the threshold amount then the ith signature bit S_i is set equal to 1, as shown in step 406. On the other hand, if the foregoing difference is greater than or equal to the arbitrary threshold value, then the ith signature bit S_i is set equal to 0, as shown in step 407.

From each of steps 406 and 407, the ith signature bit is concatenated with the previously created bits of the signature to form an updated version of the signature S, as shown in step 408. In step 409, if i is less than 6 (which is equal to the number of blocks (4 luminance plus 1 of each of two chrominance) in the macro block in the present invention), then i is incremented by 1 in step 410, and DCi and DCi' are calculated for the new value of i.

On the other hand, if i is greater than or equal to 6 in step 409, then in step 411, the absolute value of the motion vector M_{wj} is compared to 0 to determine whether the motion vector is zero or non-zero. If the absolute value of the motion vector M_{wj} is not 0, then motion bit mj is set equal to 1, in step 412. On the other hand, if the absolute value of the motion vector M_{wj} is equal to 0, then motion bit mj is set equal to 0, in step 413. Then, in step 414, a new value of signature S is further formed by concatenating the signature S with the value of the motion bit mj.

In step 415, the absolute value of the motion vector M_{wj}' is compared to 0. If the absolute value of the motion vector M_{wj}' is greater than 0, then in step 416, the motion bit mj' is set equal to one. However, if the absolute value of the motion vector M_{wj}'

is equal to 0, in step 417, the value of the motion bit mj' is set equal to 0. In step 418, the signature S and the value of motion vector mj' are concatenated to form a new value for the signature S.

In step 419, if the value of the window index j is less than the number of window pairs in the frame of the video clip, then the value of j is incremented by one in step 421 and the signature extraction process of the steps 402 through 420 is repeated. However, if the value of index j is greater than or equal to the number of windows in the frame of the video clip, then the signature extraction from one frame of the video clip is complete.

Video clips may be retrieved using their signatures. Since each video clip is represented as a sequence of signatures, these signatures are used for comparing the video clips by the Hamming distance measure.

A goal in video retrieval is to identify a query video clip, extract the signature of the query video clip, then retrieve other video clips from either local or remote, or a combination thereof, databases which are similar to the query video clip as based upon the signatures of the video clips stored in the databases.

The first step in the retrieval process is to extract the signatures from the query video clip. Then, the signature of the query video clip is compared to the signatures representing video clips stored in the database. The matching is done by measuring the conventional Hamming distance. For example, the Hamming distance measure between 0101 and 1011 is 3 because the number of different bits between 0101 and 1011 is 3.

The Hamming distance between two frames is calculated by a computer as a sum of bits set to "1" obtained in the result of the bitwise "exclusive-OR" operation.

The Hamming distance calculation provides segments of video clips from the entire database with a minimum distance, to the query video clip.

For video clips encoded using the MPEG encoding standard, motion information is also used in the retrieval process. If the coding pattern involves only I frames, as in the case of video clips encoded using the Motion JPEG encoding standard, then the matching between the query signature and the database signatures is performed by the computer using DC signatures.

The similarity of frames of the video clips is examined, with the ordering of the frames in the sequence being preserved. For example, in the present invention, the signature of a first representative frame of a query video clip is compared to the signature of a first representative frame of a database video clip. Likewise, the signature of a second representative frame of the query video clip is compared to the signature of a representative

frame of the database video clip, if there is correspondence between the second representative frame of the query video clip and the foregoing representative frame of the database video clip.

Correspondence between signatures of representative frames from the query video clip and the database video clip occurs if the database video clip representative frame from which the database video clip frame signature is extracted is at a same frame location (or within one frame thereof) in the database video clip as the query video clip representative frame from which the query video clip frame signatures is extracted, taking into account an offset number of frames between the frames in the query video clip and the database video clip.

The offsets are discussed herein with reference to Figures 11(A) through 11(D).

Subsequently, the frame signatures are shifted by one frame signature, with the first frame signature from the query video clip being compared with the second frame signature from the database video clip, etc.

The foregoing process of shifting the comparison of the frame signatures by one frame signature is repeated until the signature of the query video clip sequence has been compared against all sequences of frame signatures in the database. The similarity expected, between the query video clip and the video clips stored in the video clip database, is calculated by subtracting from 128 the total Hamming distance between corresponding frame signatures in the present invention. The highest similarity score, therefore, for each database video clip is stored.

An example of retrieval of the video clips in the present invention is discussed with reference to the following pseudo-code, That calculates the Hamming distance measure between the signature of a query video clip and the signature of a database video clip.

Assume that the signature sequence Q for the query clip is $\{q_1, \ldots, q_n\}$, and the signature sequence D for the database clip is $\{d_1, \ldots, d_m\}$. The following pseudo-code provides the Hamming distance scores, between the query clip signature and the database clip signatures.

Given Q = {q_1, . . ., q_n}, Q < > empty set

for s = 1, . . . , NUMBER_OF_VIDEO_CLIPS_IN_DATABASE

TempQ = Q

D_s = {d_1, . . ., d_m}, m is the number of representative frames in D_s,

D_s < > empty set

minscore = length of frame signature + 1

for j = 1, m-n

sum_j = 0

count_j = 0

for i = 1, n

K <- closest Frames (D, q_i)

/* K can have at most three elements */

$$\text{sum\_j} = \text{sum\_j} + \sum_{k \in K} \text{Hamming}(\text{Sig}(q\_i), \text{Sig}(d\_k))$$

count_j = count_j + | K |

endfor i

score_j = sum_j / count_j

if (score_j < minscore), then minscore = score_j

TempQ = { q_p | q_p <- q_p + d_j+1 - d_j, for p = 1,n}

endfor j

Scores <-- insert (minscore, s)

endfor s

The total Hamming distance between the signature of a query video clip and the signature of a database video clip is calculated for all database video clips. The signatures correspond to frames in each video clip, but the distance between the frames for which respective signatures were extracted in the query sequence and between the frames for which the respective signatures were extracted in each database sequence are not necessarily equal; the distance can be selected arbitrarily. For example, a frame signature may have been extracted for each of frames 1, 2, and 7 of the query video clip, but for each of frames 1, 5, 11, and 15 of the database video clip if each of the above-mentioned frames are representative of the video clip. On the other hand, each fifth (or second, or third, etc., for example) frame may be arbitrarily selected for signature extraction in the video clip.

In the above-mentioned pseudo-code, Q is a set containing "n" entries and comprises the signature sequence extracted from the query video clip. D_s is a set containing "m" entries and comprises the signature sequence extracted from the database video clip "s". The "for" j loop is repeated, for each database video clip "s" stored in the video clip signature sequence database.

At the start of comparison between the query video clip signature sequence and each database video clip signature sequence, to preserve query video clip signature sequence Q, TempQ is initialized to Q. TempQ is then manipulated during the comparison between the query video clip signature sequence and the database video clip signature sequence "s". Then, the variable minscore is initialized to the value of the length of the frame signature + 1.

In the pseudo-code, an index j determining the number of iterations of comparison between the signature of the query video clip and the signature of the database video clip is initialized. The index j is based on the number of frame signatures in each of the foregoing video clips, as shown. Variables sum_j, which indicates the Hamming distance for the comparison of the signature of the query video clip to the signature of the database video clip for iteration j, and count_j, which is the number of comparisons between frame signatures for the foregoing video clips, are also initialized. An index i indicates the number

of the representative frame signature in the query video clip signature sequence. An index k indicates the number of frames in the database video clip which are within no more than one frame of the representative frame in the query video clip, taking into account any shifting or offset of frames.

In the pseudo-code, each frame signature of the query video clip signature is compared to a corresponding frame signature of the database video clip signature. In addition, each frame signature of the query video clip signature is compared to the previous or next frame signature of the database video clip signature, if correspondence (as explained in detail above) exists. Accordingly, each representative frame, for which a signature has been extracted, of the query video clip is compared to each representative frame of the database sequence video clip corresponding to or within one frame (which is preferred but which could be any pre-established number) in either frame direction of the corresponding frame from the query video clip, taking into account the relative starting positions between the frames of the query sequence and the frames of the video clips.

For each of the one, two, or three (which value is indicated by "k") frames meeting the above-mentioned criteria from the query video clip frame (as explained herein above), the Hamming distance between the signature of the video clip and the signature of the database clip for the frames being compared is determined, and sum_j and count_j are updated accordingly.

In the foregoing pseudo-code, the Hamming distance is computed for each of the one, two, or three frames which are within one frame of the query video clip frame.

After the comparison between the signature of the query video clip and the signature of the database video clip is complete for the current offset between the entries of the query video clip signature sequence and the database video clip signature sequence, the average (score_j) is calculated by dividing the sum of the Hamming distances between the signature of a query representative frame and the signatures of the representative, corresponding frames by the number of Hamming distances calculated.

Since minscore stores the value of the lowest Hamming distance measure between the signature sequence extracted from the query video clip and the signature sequence extracted from the database video clip, if the Hamming distance measure calculated for the current offset between the entries in the query video clip signature sequence and the database video clip signature sequence is lower than the Hamming distance measure calculated for the same database video clip signature sequence at any prior offset, the value of minscore is replaced by the value of score_j.

Since the number of frames in the database video clip is assumed larger than the number of frames in the query video clip, the comparison between the signature of the query video clip and the database video clip is repeated a number of times equal to a number of sequences of frame signatures in the database video clip in which the sequence of frame signatures in the query video clip may exist, preserving the sequence of frame signatures within each video clip (i.e., keeping the video clip frame signatures in the same order). Index j determines when the foregoing criteria has been met.

For each subsequent comparison between the query video clip signature and the database video clip signature, the query video clip signature is offset against the database video clip signature by one representative frame. In the above-mentioned pseudo-code,

$$TempQ = \{ q\_p \mid q\_p <- q\_p + d\_j+1 - d\_j, \text{ for } p = 1,n\}$$

offsets the entries in the query video clip sequence so that the signature extracted from the first representative frame in the query video clip is offset to correspond to the next entry in the database video clip signature sequence. The next entry in the database video clip signature sequence is the entry in the database video clip signature sequence immediately subsequent to the entry in the database video clip signature sequence to which the signature extracted from the first representative frame in the query video clip signature sequence corresponded in the previous iteration of the "for" j loop.

At the completion of the "for" j loop for the current database video clip signature sequence "s" in the database, the value of minscore and the corresponding, current value of "s" are inserted into an array Scores. Scores stores the lowest value of the Hamming distance measure for each database video clip against which the query video clip signature sequence was compared in accordance with the present invention.

The array Scores is, in a preferred embodiment, a linked list, which is sorted based on the lower scores. The lowest score stored in Scores indicates the best "match" between the query video clip signature sequence and the database video clip signature sequence.

If similarity between the signature of the query video clip and the signature of each, respective database video clip is used as the basis for sorting database video clips, then the database video clips are arranged in descending order of similarity.

With the above-mentioned pseudo-code of the present invention, the search can be started with any signature in the database.

In a preferred embodiment, when calculating the measure of the distance

between the signatures of two video clips, the average of all scores of frame signatures compared is used. However, the similarity measure may be based upon other criteria, such as local resemblance between two video clips. The following are methods at the frame level for determining video similarity measures in the present invention:

1. Using the average overall scores of matched frames (which is shown in the above-mentioned pseudo-code);
2. Using the average over part of the highest score of matched frames; or
3. Using the average over the local scores in the proximity of a few local maximae, in which the averages of the similarity between signatures of representative frames providing the highest similarity are used.

For the present invention also methods at levels other than the frame level may also be employed.

The above-mentioned similarity measures provide the overall similarity between two video clips, taking into account the linear appearance of spatial and motion features of the respective video clips. These methods are useful when the similarity between two video clips should be relative to the linear placement of the scenes in the respective video clips. However, when the overall similarity of subsegments of two video clips is needed, regardless of the placement of the scenes in the video (for example, the beginning frames of the query video clip can be matched with the ending frames of the database video clip and vice-versa), the average of the highest scores of the similar common subsegments must be used.

Further, in the present invention, users are able to attach importance to various parts of the video query clip, to various aspects of video representation such as spatial and motion aspects. The user is able to:

1. Attach importance to a subsequence of frames;
2. Attach importance to features such as luminance, chrominance, and motion components of video clips;
3. Select the density of representative frames in the sequence of the representative frames to be used in the searching process; and
4. Select video frames which are important in the searching process.

An example of a comparison between signature of the query video clip and the signature of a database video clip consistent with the above-mentioned pseudo-code is

shown in Figures 11(A) through 11(D). In Figures 11(A) through 11(D), the value of m=13 and n=4; therefore, m-n=4 iterations of comparisons between the above-mentioned signature sequences are made.

In the retrieval process, the signatures of frames from the query video clip are compared to the signatures of frames from each database video clip. Figures 11(A) through 11(D) indicate by "R" which frame numbers from each of the query video clip and the database video clip are compared. Frames being compared in accordance with the aforementioned pseudo-code and criteria are indicated by arrows.

In Figure 11(A), representative frames, which are frames for which signatures have been extracted are designated by "R" for each of the query video clip and the database video clip. The representative frames can be placed at any frame position, either randomly or at regular intervals, or at each frame. Even the representative frames may be placed in an irregular pattern.

Also as shown in Figure 11(A), frame 1 of the query video clip is compared with frame 1 of the database video clip. Then, since the next frame of the database video clip having a signature is frame 4, the distance between frame 1 of the query video clip and frame 4 of the database video clip is too large (i.e., exceeds the distances of one frame arbitrarily selected). Therefore, frame 1 of the query video clip is compared solely with frame 1 of the database video clip.

The distance for correspondence between the frame number of the representative frame of the query video clip and the frame number of the representative database frame from the database video clip is selected arbitrarily; in the case of Figures 11(A) through 11(D) example, that parameter is chosen arbitrarily as one. Therefore, if a representative frame in the query video clip is frame number 6, then the signature extracted from frame number 6 of the query video clip is compared only with signatures which exist for frame numbers 5, 6, or 7 from the database video clip. Accordingly, in the example of Figure 11(A), even though a signature for frame number 4 of the database video clip exists, there is no frame from the query video clip which is eligible to be compared with the signature of frame number 4 from the database video clip, based on the chosen parameter of a distance of one frame.

In Figure 11(A), frame 6 of the query video clip is a representative frame the signature for which is compared with the signature for frame 7 from the database video clip, also a representative frame within one frame of frame 6 of the query video clip. Even though, as shown in Figure 11(A), frame 8 of the query video clip is a representative frame,

and frame 7 of the database video clip, also a representative frame, and is within one frame of frame 8 from the query video clip, frame 8 from the query video clip is not compared with frame 7 of the database video clip since frame 7 of the database video clip has already been compared with frame 6 from the query video clip.

In the present invention, once a signature from a representative frame from the database video clip has been compared with a signature from a representative frame from the query video clip, that signature of the representative frame from the database video clip is not compared with a signature from another representative frame from the query video clip until the entries in the query video clip signature sequence are offset against the entries in the database video clip signature sequence. However, this restriction of not using a signature from a representative frame from the database video clip for a second comparison may be removed.

Figure 11(A) also shows that the frame signatures from query video clip frames 11, 13, 15, and 17 are compared with frame signatures from representative database video clip frames in accordance with the above-mentioned description. Respective signatures from representative frames 20 and 24 from the query video clip are each compared with respective frame signatures from two frames from the database video clip; each of the foregoing two database video clip representative frames falls within one frame of the corresponding representative query video clip frame. Therefore, two "scores" for each of representative frames 20 and 24 from the query video clip are obtained. In that case, the two respective "scores" are averaged for each of representative video clip frame 20 and representative query video clip 24, as indicated in the above-mentioned pseudo-code.

In Figure 11(B), the same query video clip as shown in Figure 11(A), is off-shifted to the second representative frame signature from the database video clip, in accordance with the pseudo-code described herein above.

As shown in Figure 11(B), the signature extracted from representative frame 1 from the query video clip is compared with the signature extracted from representative frame 4 from the database video clip. Then, since a second representative frame from the database video clip is not within one frame of the representative frame from the query video clip, taking into account the off-set positions of the two respective video clips, frame 1 from the query video clip is not compared with any other frames from the database video clip. Likewise, representative frame 6 from the query video clip does not correspond to a representative frame from the database video clip, nor is within one frame of the database video clip taking into account the foregoing offset position. Therefore, the

signature extracted from representative frame 6 from the query video clip is not compared with the signatures extracted from any frames from the database video clip.

In Figure 11(B), the signature extracted from representative frame 8 from the query video clip is compared with the signature extracted from representative frame 11 from the database video clip. Representative frame 8 from the query video clip corresponds to representative frame 11 from the database video clip, taking into account the off-set between the query video clip frames and the database video clips frames. Likewise, the signature extracted from representative frame 11 from the query video clip is compared with the signature extracted from representative frame 14 from the database video clip. The respective signatures extracted from representative frames 13, 15, 17, 20, and 24 from the query video clip are compared with respective signatures extracted from corresponding representative frames from the database video clip.

In Figure 11(C), the signature extracted from representative frame 1 from the query video clip is compared with the signature extracted from representative frame 7 from the database video clip, representative frame 7 being the next representative frame to which representative frame 1 from the query video clip is offset. Likewise, the signature extracted from representative frame 6 from the query video clip is compared with the signature extracted from representative frame 11 from the database video clip, etc.

In Figure 11(D), the signature extracted from representative frame 1 from the query video clip is compared with the signature extracted from representative frame 11 from the database video clip, representative frame 11 being the next representative frame to which representative frame 1 from the query video clip is offset. Likewise, the signature extracted from representative frame 6 from the query video clip is compared with the respective signatures extracted from representative frames 15 and 17 from the database video clip, etc.

In the examples in Figures 11(A) through 11(D), the difference in the number of frames between the first representative frame from the query video clip and the representative frame from the database video clip to which the representative frame from the query video clip is compared is preserved for each frame signature comparison between the query video clip and the database video clip.

Figure 12 shows an example of an implementation of a user interface 32 described herein above with reference to Figure 3. The user interface 46 shown in Figure 12 is implemented using a TCL/TK toolkit running on an X WINDOWS™ platform. When a user selects "Search by Video" 48, a topic such as "NEWS" 50, and a query video clip 52

(which is shown as the upper leftmost icon in the workspace for retrieval area 54), then the present invention searches extracted signatures from a meta database 22 (described herein above with reference to Figure 3) and displays results of that search in the workspace for retrieval area 54. The resultant video clips are ordered according to similarity level to the query video clip, from left to right across rows, then from top to bottom of columns. Screen window 56 displays a thumbnail (or reduced image) of a selected video.

The present invention also encompasses variations of the embodiments.

The many features and advantages of the invention are apparent from the detailed specification and, thus, it is intended by the appended claims to cover all such features and advantages of the invention which fall within the true spirit and scope of the invention. Further, since numerous modifications and changes will readily occur to those skilled in the art, it is not desired to limit the invention to the exact construction and operation illustrated and described, and accordingly all suitable modifications and equivalents may be resorted to, falling within the scope of the invention.

Claims:

1. A method of a computer for identifying video clips similar to a query video clip, said method comprising:

extracting a signature of each of the video clips and the query video clip;

determining a score of the similarity of the signature of each of the video clips to the signature of the query video clip; and

ordering the video clips based on the score.

2. The method according to Claim 1, wherein the score is determined by calculating the Hamming distance measure between the signatures.

3. The method according to Claims 1 or 2, further comprising displaying the video clips based on the score.

4. The method according to Claims 1, 2 or 3, further comprising the step of digitally encoding the video clips into the MPEG format, if the video clips are not encoded in the MPEG format.

5. A method comprising the steps of:

extracting a signature of a query video clip;

comparing the signature of the query video clip to signatures of the video clips in a database; and

retrieving a video clip responsive to the comparison.

6. The method according to Claim 5, wherein the database stores the location, size, and time length of the video clips corresponding to the signatures.

7. An apparatus comprising:

a source supplying video clips; and

an information retrieval system extracting signatures from the video clips and the signature from a query video clip, identifying video clips similar to the query video clip by comparing the signature of the query video clip to the signatures of the video clips, and retrieving from the video source the video clips similar to the query video clip.

8. An apparatus comprising:

a source supplying video clips and video clip signatures; and

an information retrieval system extracting a query signature from a query video

clip, comparing the query signature to the video clip signatures of the video clips, and retrieving from the video source a video clip responsive to the comparison.

9. The apparatus according to Claim 8, wherein the video information retrieval system further comprises a retrieval subsystem determining similarity between the signature of the query video clip to the signatures of the video clips.

10. The apparatus according to Claims 8 or 9, further comprising a database storing the signatures.

11. The apparatus according to Claims 8, 9 or 10, wherein the database stores the location, size, and time length of the video clips corresponding to each of the signatures.

12. A computer retrieving and displaying video clips from video sources and similar to a query video clip, said apparatus comprising:

a source supplying video clips;

an information retrieval system extracting signatures from the video clips and the signature from a query video clip, identifying video clips similar to the query video clip by comparing the signature of the query video clip to the signatures of the video clips, and retrieving from the source the video clips similar to the query video clip; and

a display displaying the video clips similar to the query video clip.

13. An apparatus retrieving and displaying video clips from video sources and similar to a query video clip, said apparatus comprising:

a source supplying video clips;

an information retrieval system comprising:

an archival and signature extraction section encoding each of the query video clip and the video clips into an MPEG format if the query video clip and the video clips are not in the MPEG format, and extracting signatures of query video clip and the video clips,

a database storing the signatures of the video clip, and

a retrieval subsystem identifying the video clips similar to the query video clip by comparing the signature of the query video clip to the signatures of the video clips; and

a display displaying the video clips similar to the query video clip.

14. The apparatus according to Claim 13, wherein the database stores the location, size, and time length of the video clips corresponding to each of the signatures.

15. The method according to Claim 6, wherein similarity between the

signatures of the video clips and the signature of the query video clip is determined by Hamming distance measure.

16. The method according to Claim 15, wherein the Hamming distance measure is determined between each signature frame of the query video clip and three or fewer signature frames of one of the database video clips.

17. The method according to Claim 1, further comprising the step of digitally encoding the video clips into the Motion JPEG format, if the video clips are not encoded in the Motion JPEG format.

18. An apparatus comprising:

a video source providing video clips, said video source comprising:

an MPEG video server providing the video clips encoded using an MPEG encoding standard,

a live video source providing the video clips, and

a network source providing the video clips encoded using an encoding standard other than the MPEG encoding standard;

a video information retrieval system, coupled to the video source, extracting signatures from the video clips and comparing a signature extracted from a query video clip to the signatures, said video information retrieval system comprising:

an archival and signature extractor one of encoding and partially-re-encoding using the MPEG encoding standard the video clips if the video clips are not encoded using the MPEG encoding standard, and extracting the signatures from the video clips if the video clips are encoded using the MPEG encoding standard,

a database, coupled to the archival and signature extractor, storing the signatures, along with identifying data of the respective video clips corresponding to the signatures, and

a retrieval subsystem, coupled to the database, comparing the signature of a query video clip with the signatures stored in the database, and determining a similarity between the signature of the query video clip and each of the signature stored in the database, said retrieval subsystem comprising:

a similarity calculator calculating the similarity;

an ordering unit, coupled to the similarity calculator, ordering the signatures based on the similarity; and

a fetch and display unit, coupled to the ordering unit, for retrieving video clips corresponding to the signatures; and

a user interface, coupled to the video retrieval subsystem, displaying the video clips.

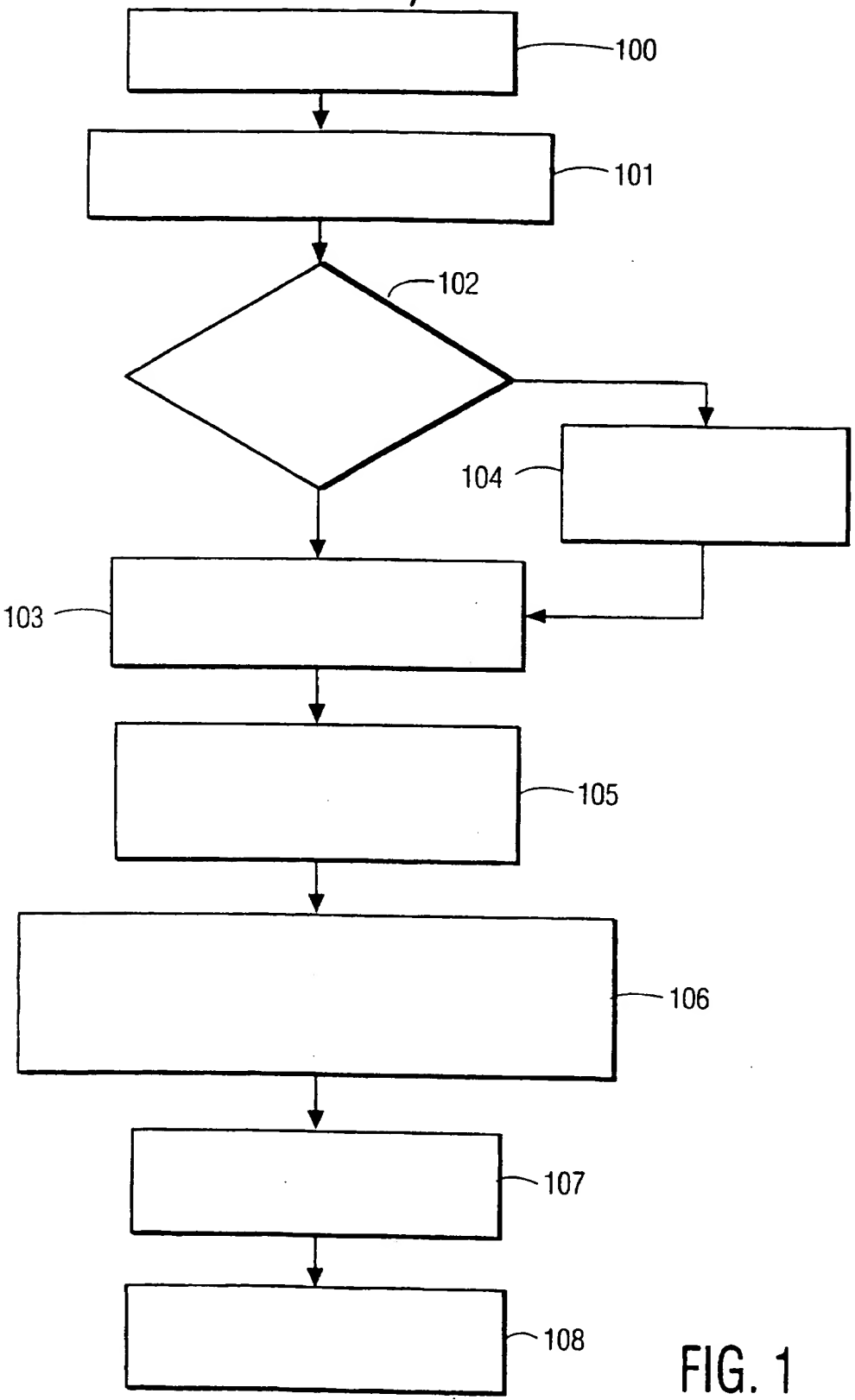


FIG. 1

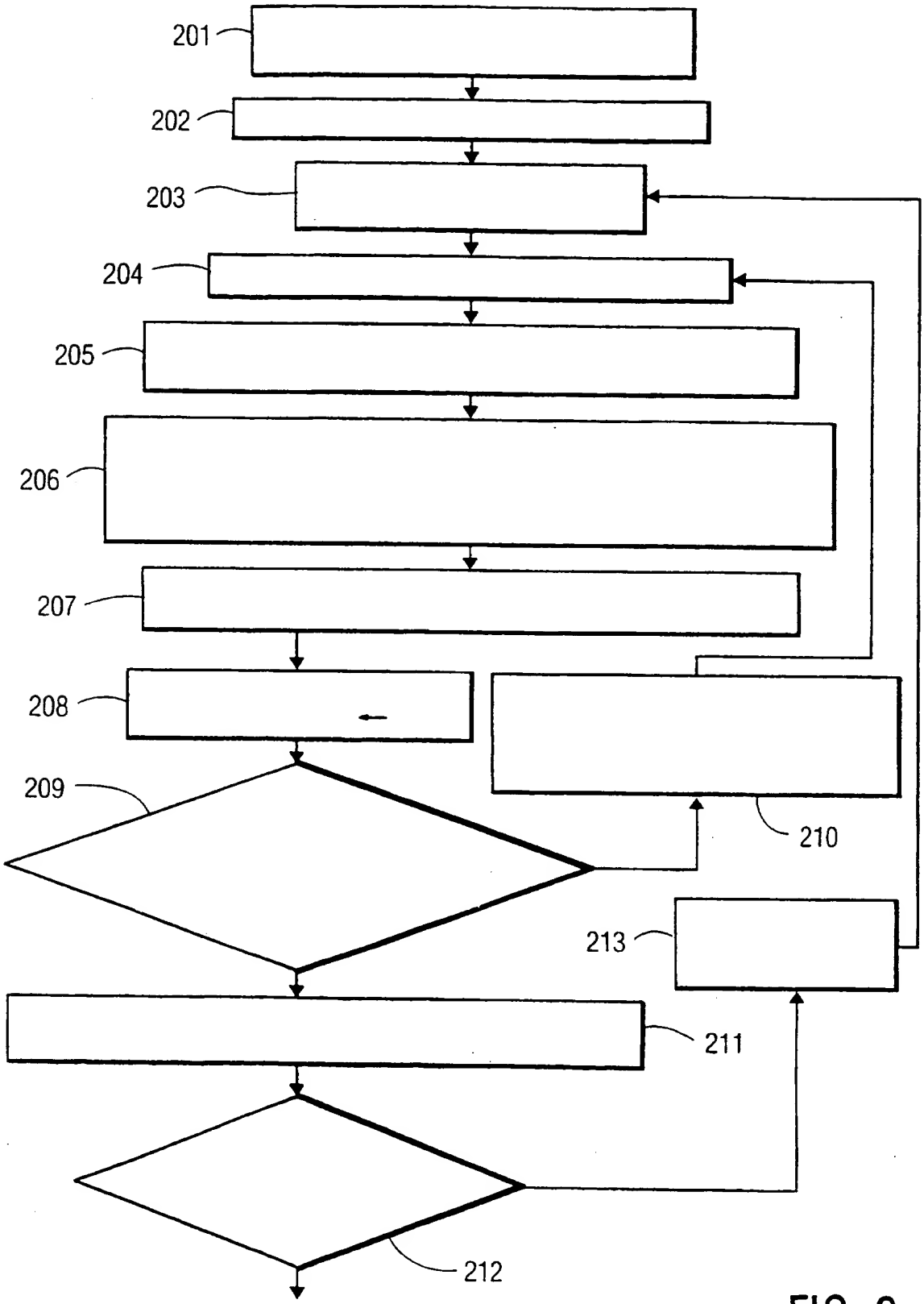


FIG. 2

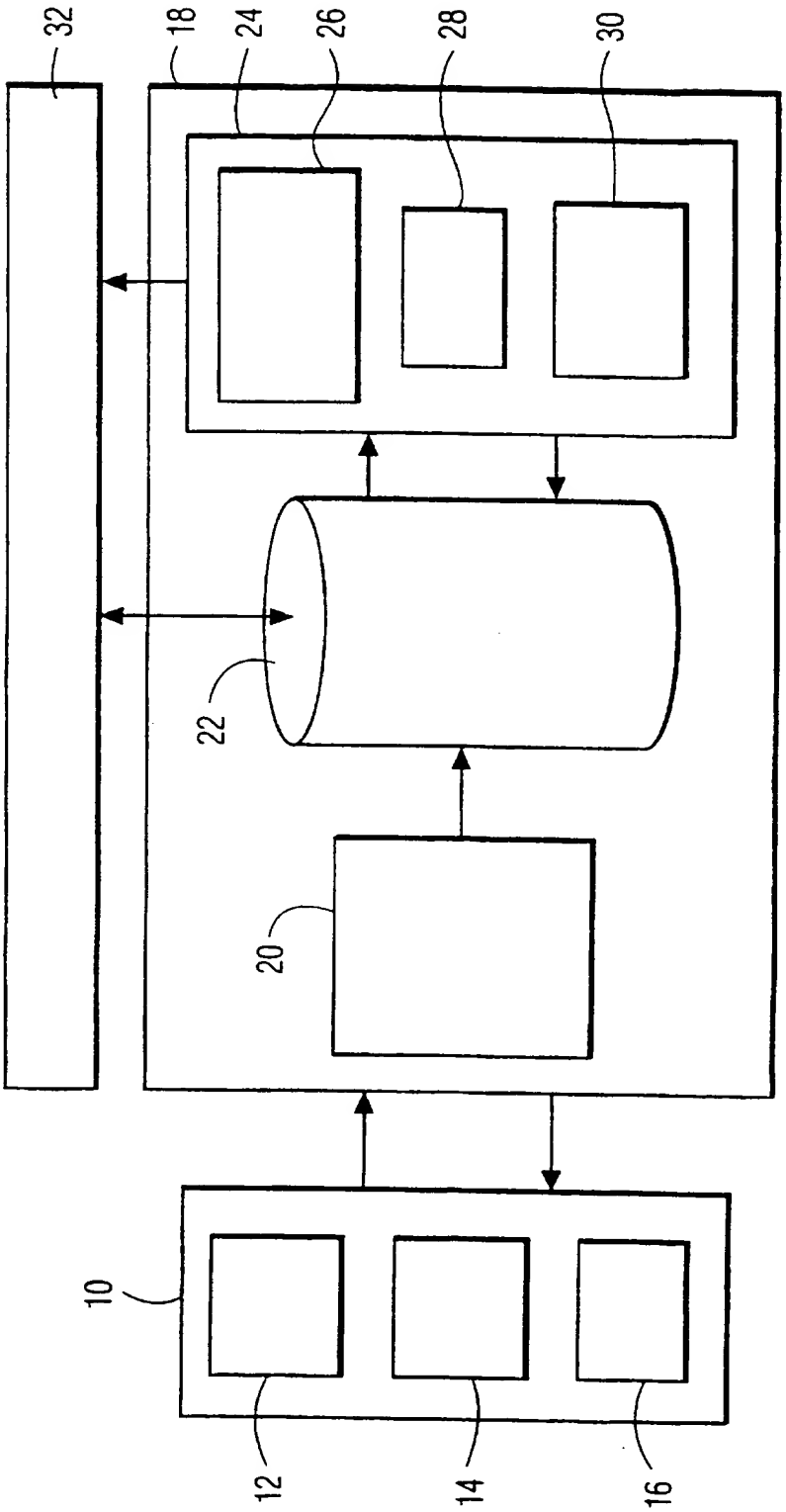


FIG. 3

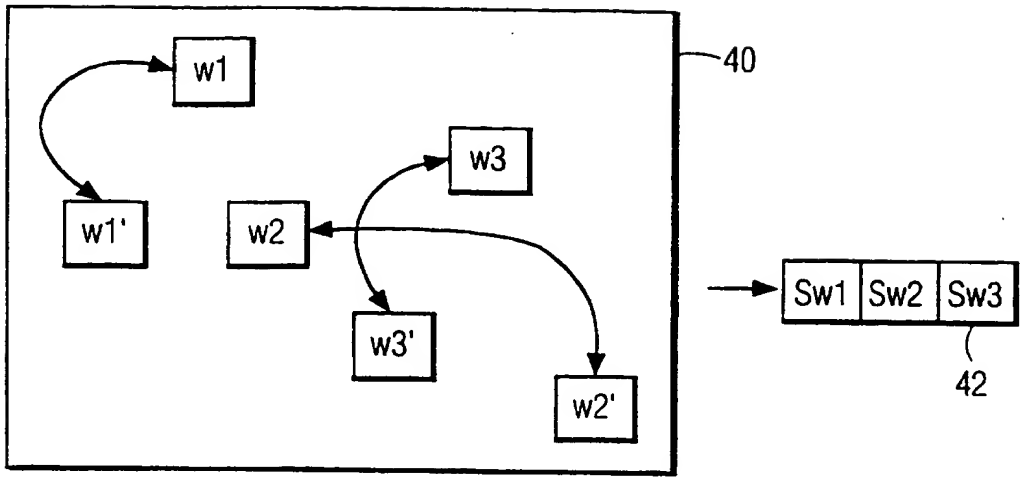


FIG. 4

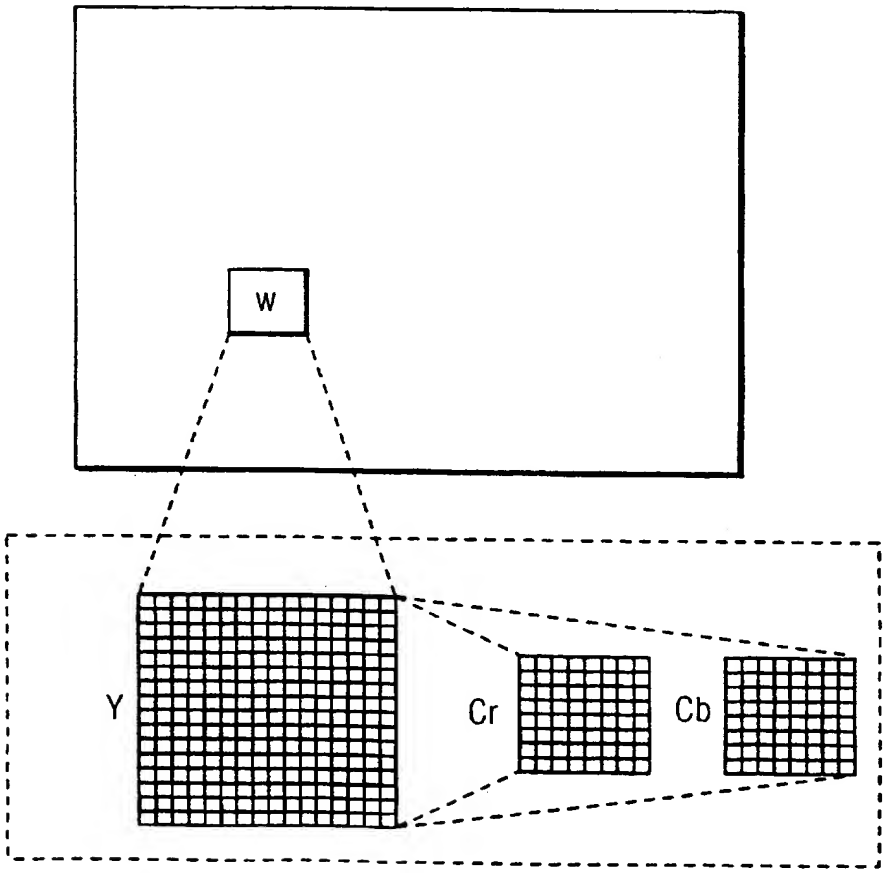


FIG. 5

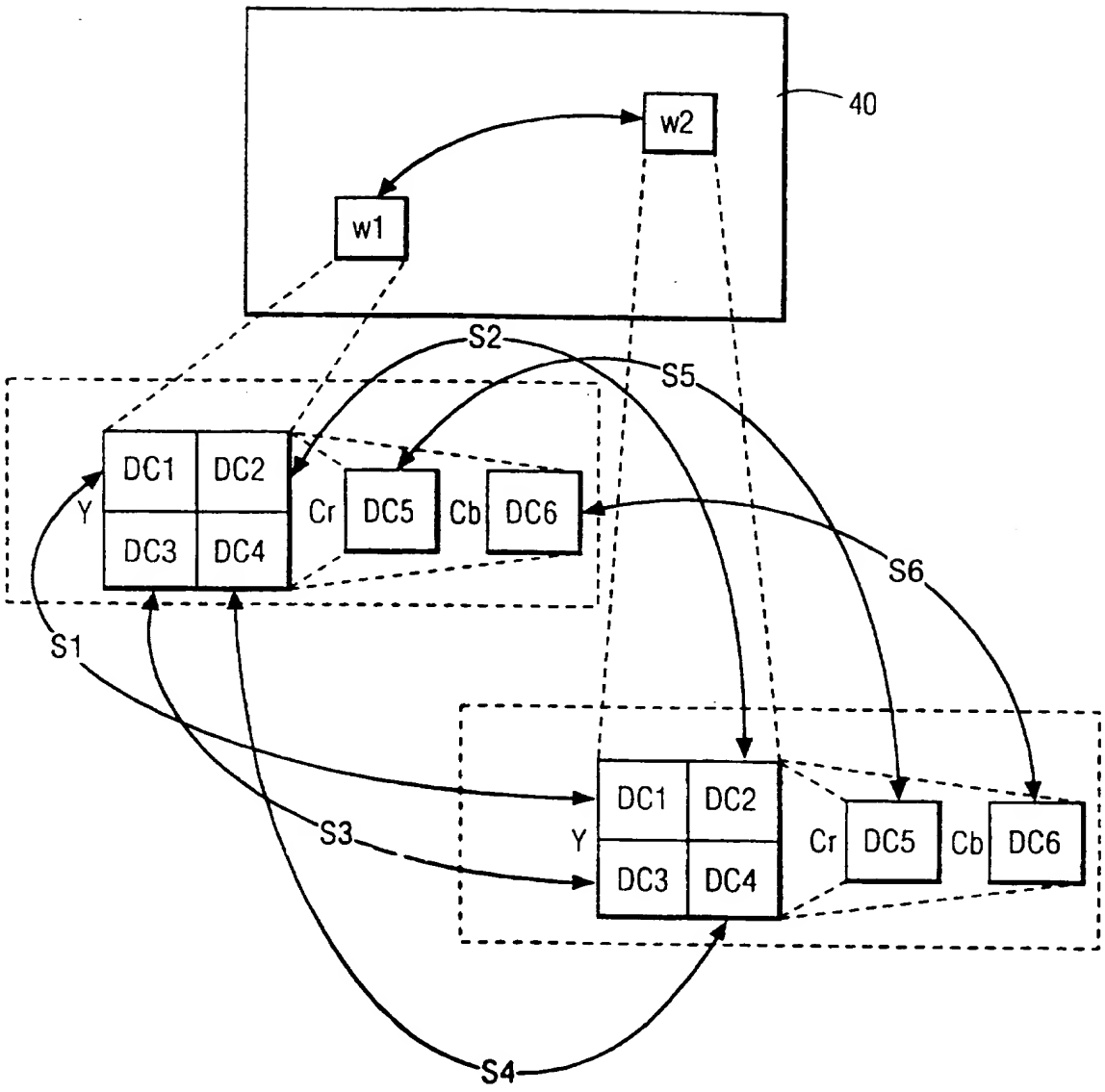


FIG. 6

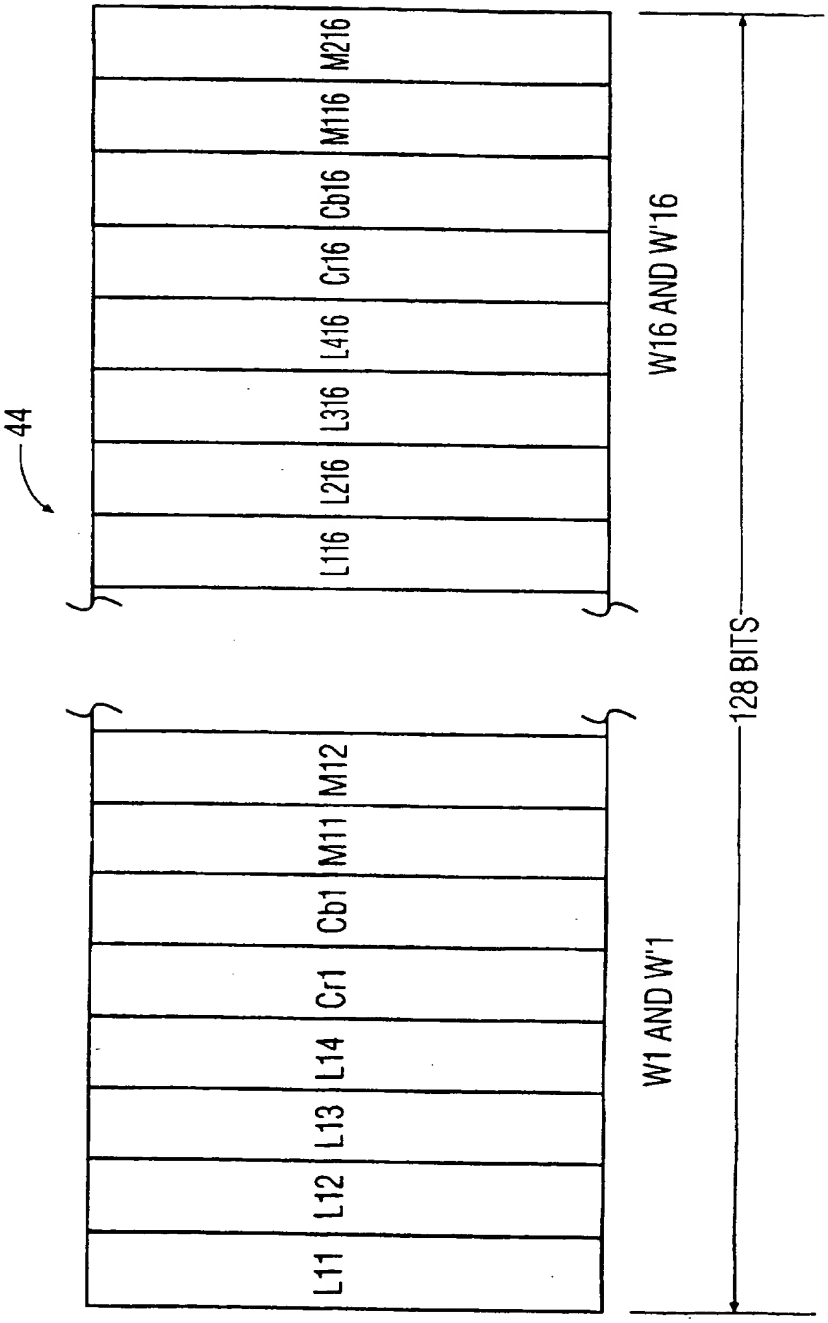


FIG. 7

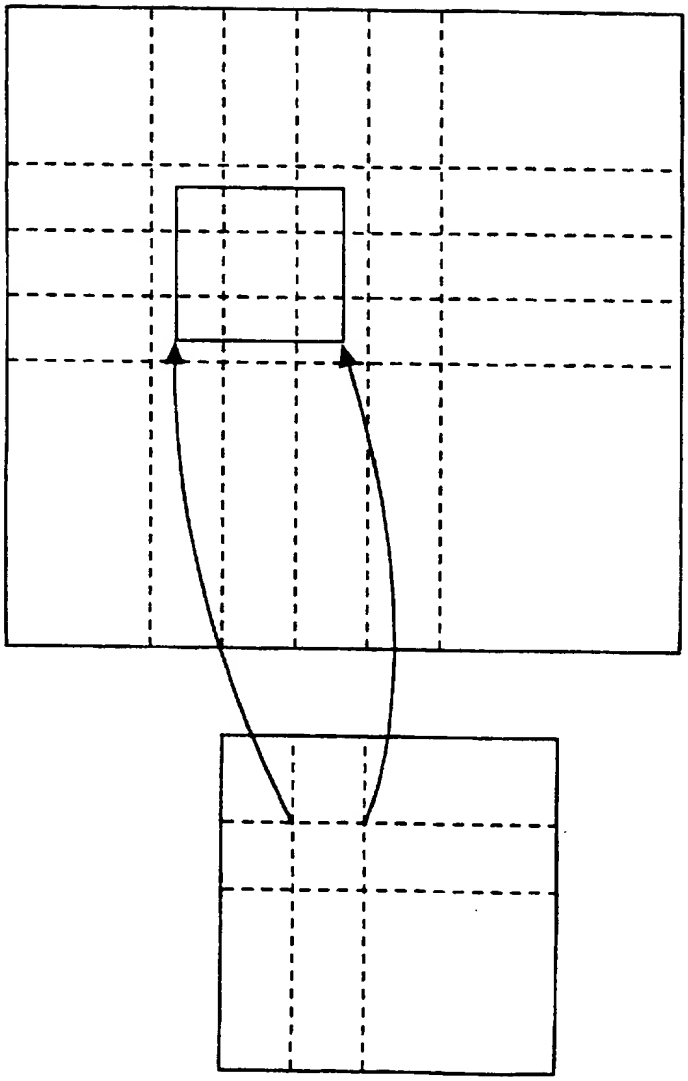


FIG. 8B

FIG. 8A

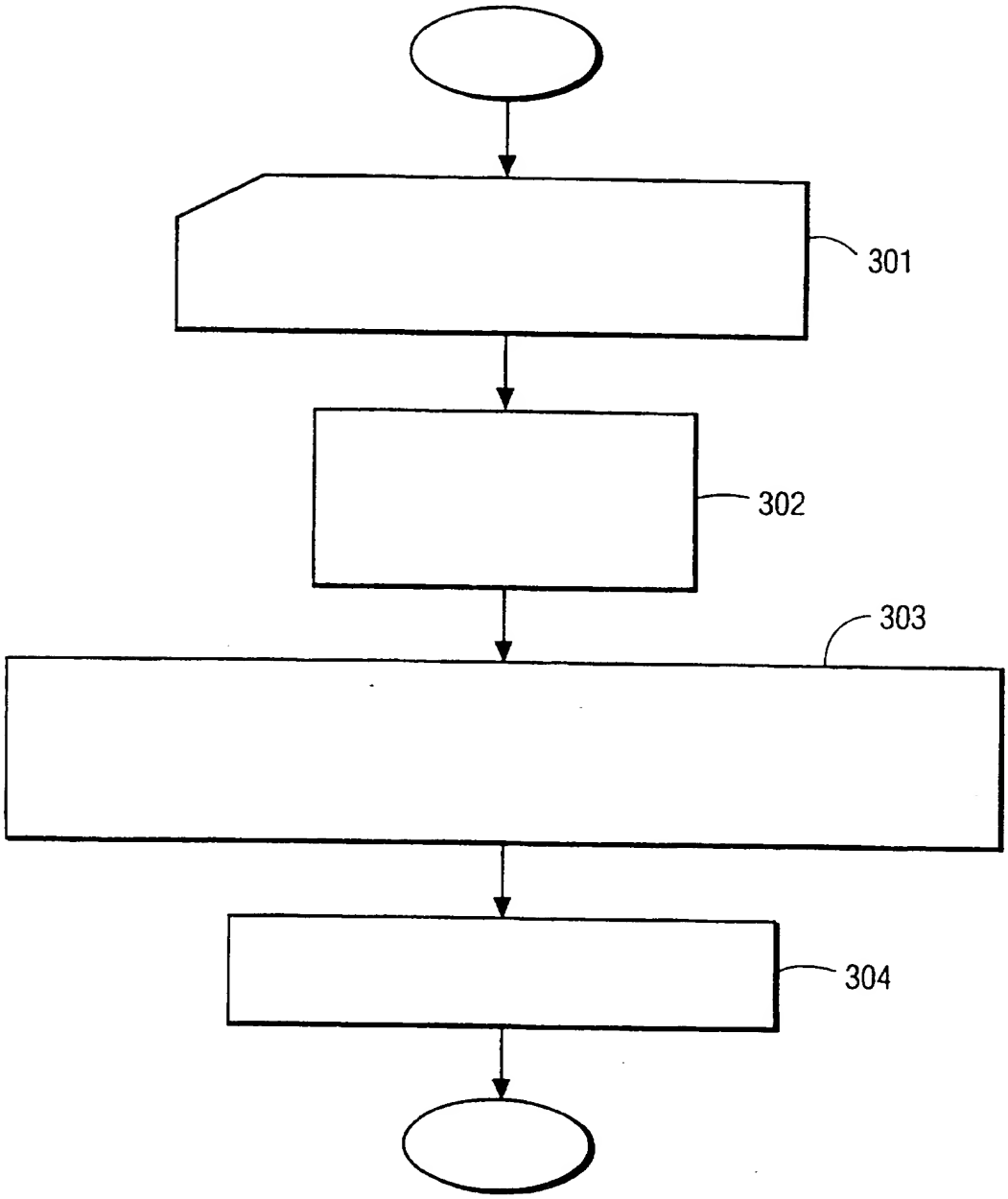


FIG. 9

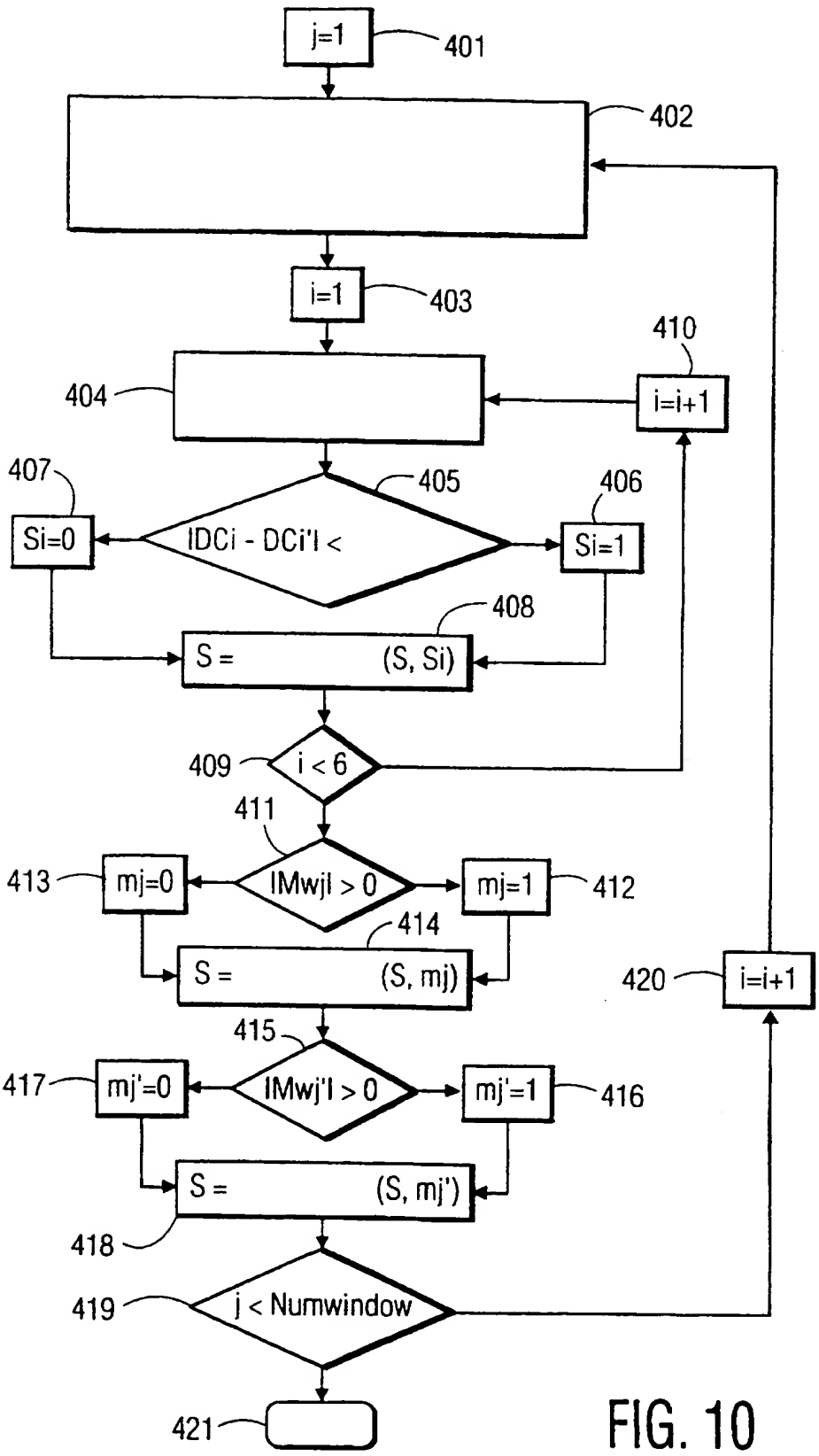


FIG. 10

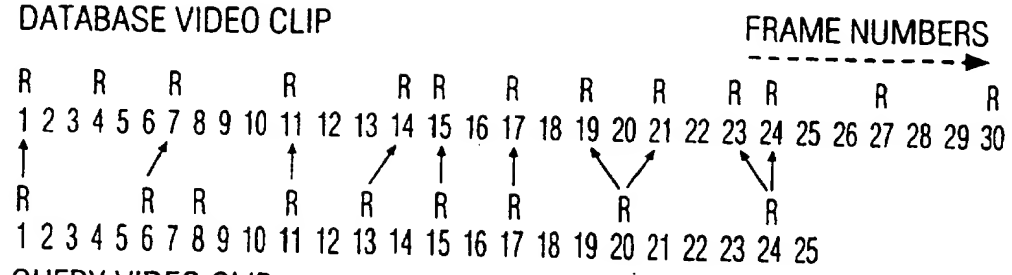


FIG. 11A

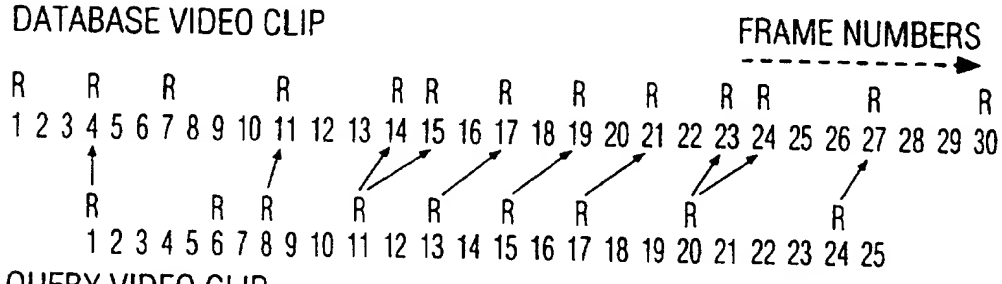


FIG. 11B

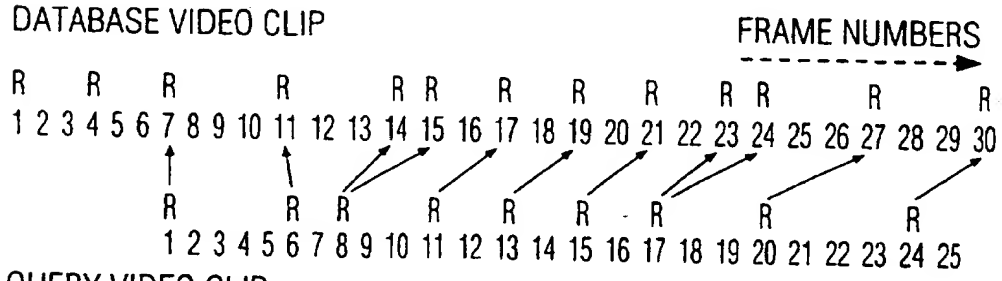


FIG. 11C

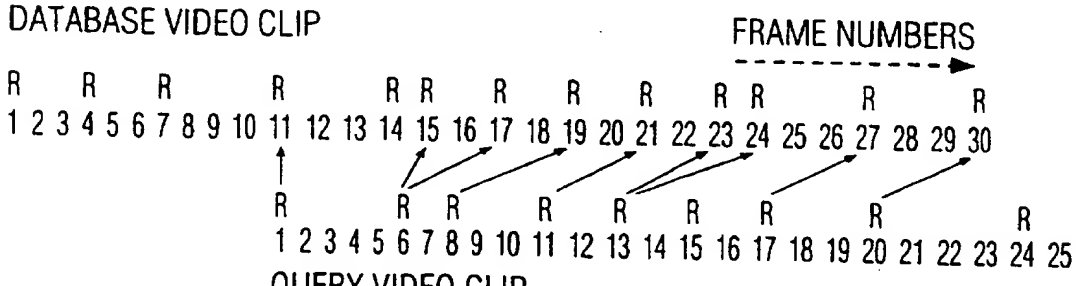


FIG. 11D

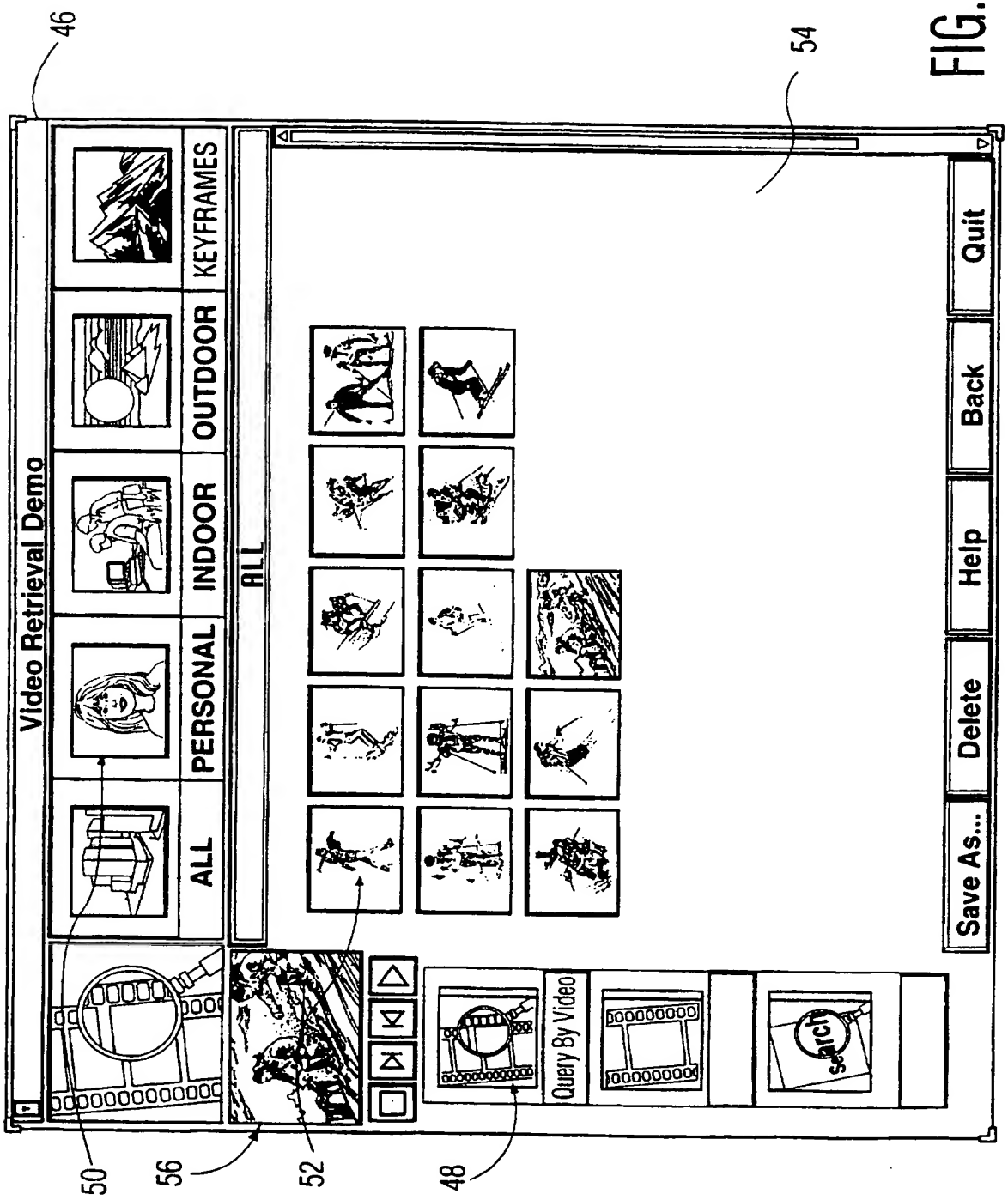


FIG. 12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB 97/00439

A. CLASSIFICATION OF SUBJECT MATTER

IPC6: G06F 17/30
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC6: G06F, G11B, H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI, PAJ, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	Proceedings of the SPIE - The International Society for Optical Engineering Conference Title: Proc. SPIE-Int. Soc. Opt. Eng. (USA), vol. 3022, p. 59-70, 1997, Dimitrova N. et al: "Content-based video retrieval by example video clip", see the whole document --	1-18
P,A	Proceedings of the SPIE - The International Society for Optical Engineering Conference Title: Proc. SPIE - Int. Soc. Opt. Eng (USA), Vol. 3022 p. 200-11 publ.date 1997, Kobla V. et al: "Compressed domain video indexing techniques using DCT and motion vector informaton in MPEG video", see the whole document --	1-18

[X] Further documents are listed in the continuation of Box C. [X] See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" erlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
2 October 1997	06 -10- 1997
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Authorized officer Irene Turcu Telephone No. +46 8 782 25 00

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB 97/00439

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,A	Proceedings of the SPIE - The International Society for Optical Engineering Conference Title: Proc. SPIE - Int. Soc. Opt. Eng. (USA), vol. 2916, p. 78-89. 1996, Kobla V. et al: "Archiving, indexing and retrieval of video in the compressed domain", see the whole document	1-18
	--	
A	EP 0643358 A2 (CANON KABUSHIKI KAISHA), 15 March 1995 (15.03.95), see the whole document	1-18
	--	
A	WO 9637074 A1 (PHILIPS ELECTRONICS N.V.), 21 November 1996 (21.11.96), see the whole document	1-18
	--	
A	EP 0675495 A2 (SIEMENS CORPORATE RESEARCH, INC.), 4 October 1995 (04.10.95), see the whole document	1-18
	--	

INTERNATIONAL SEARCH REPORT

Information on patent family members

01/09/97

International application No.
PCT/IB 97/00439

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0643358 A2	15/03/95	JP 7073195 A	17/03/95
		US 5586197 A	17/12/96
WO 9637074 A1	21/11/96	EP 0771504 A	07/05/97
		GB 9510093 D	00/00/00
EP 0675495 A2	04/10/95	NONE	

Form PCT/ISA/210 (patent family annex) (July 1992)